CRY 2025 - Laboratoire 1

25/03/2025

Maxime Lestiboudois

Questions

1. Quel est l'avantage d'utiliser le test du χ^2 plutôt que de comparer simplement la lettre la plus fréquente dans le texte chiffré par rapport aux statistiques du langage de base ?

L'avantage du χ^2 est qu'il permet d'enlever l'apparence aléatoire de la distribution des lettre. Le χ^2 permet d'analyser la distribution de toutes les lettres, ainsi que leur fréquence. Il devient ainsi un test plus robuste lors de déchiffrage de textes chiffrés.

2. Pourquoi est-ce que le test du χ^2 ne fonctionne-t-il pas directement sur un texte chiffré à l'aide du chiffre de Vigenère ?

Le χ^2 ne fonctionne pas pour un texte chiffré à l'aide du chiffre de Vigenère car l'ensemble du texte est chiffré avec un mot (clé) et non un décalage. Ainsi, il existe un nombre n, soit n la taille de la clé, de décalages présents dans le texte chiffré.

3. Que mesure l'indice de coïncidence ?

L'indice de coïncidence mesure la probabilité que deux lettres choisies au hasard soient identiques. PLus l'indice est élevé, plus il y a une répétition des lettres dans le texte. Il existe un indice de coïncidence différents pour chaque langue.

4. Pourquoi est-ce que l'indice de coïncidence n'est-il pas modifié lorsque l'on applique le chiffre de César généralisé sur un texte ?

L'indice de coïncidence n'est pas modifié lorsqu'on utilise César car la distribution des lettres n'est pas altérée. Comme l'indice de coïncidence dépend de la fréquence des lettres, la structure statistique d'un texte chiffré avec le chiffre de César est conservée.

5. Est-il possible de coder un logiciel permettant de décrypter un document chiffré avec le chiffre de Vigenère et une clef ayant la même taille que le texte clair ? Justifiez.

Oui c'est possible. C'est une opération qui se rapprocherait du chiffrement One-Time-Pad. Déchiffrer un tel texte en connaissant la clé est facile, il s'agit d'un vigenère classique, cependant déchiffrer le texte en "cassant" la clé est infiniment plus compliqué. En effet, pour casser le chiffre de Vigenère, nous nous basions sur l'indice de coïncidence calculé à l'aide des ièmes lettres (soit i la taille de la clé) du texte, ce qui nous permettait d'avoir des textes avec des décalages de lettres constants. Sans cette répétition, il est beaucoup plus compliqué de trouver la clé de déchiffrement, mais ce n'est cependant pas impossible.

6. Expliquez votre attaque sur la version améliorée du chiffre de Vigenère.

Nous avons pu remarqué qu'il est possible de récupérer un texte chiffré avec le chiffre de Vigenère en connaissant la taille de la clé, soit le nombre d'éléments dans un bloc. Pour trouver le contenu d'un bloc n chiffré avec le chiffre de Vigenère, il suffit de lui "soustraire" tous les n-i èmes, soit i de 0 à n-1, (application du déchiffrement du chiffre de Vigenère d'un bloc par ses précédents blocs).

En ayant connaissance de cela, nous pouvons casser la version améliorée du chiffre de Vigenère:

1. Récolter tous les textes possibles de chiffrés avec le chiffre de Vigenère, soit décoder le texte selon la méthode vu ci-dessus pour chaque taille de clé entre 1 et 20.
2. Pour chacun des textes trouvés précédemment, trouver la clé en utilisant la fonction vigenre_break() créée auparavant.
3. Déchiffrer les différents textes chiffrés avec le chiffre de Vigenère avec leur clé respective (fonction vigenere_decrypt()).
4. Pour chacun des textes déchiffrés au point 3, calculer leur indice de coïncidence et les stocker dans un tableau.
5. Trouver l'indice de coïncidence des différents textes le plus proche de l'indice de coïncidence de référence. Son indice dans le tableau + 1 est la longueur de la clé. Soit L la longueur de la clé
6. Créer les différents textes prenant chacunes des lettres d'indice i mod L, i entre 0 et la longueur du texte (non comprise).
7. Trouver la clé à l'aide du déchiffrement du chiffre de César sur les différents textes trouvés au point 6.

7. Trouvez une méthode statistique (proche de ce qu'on a vu dans ce labo) permettant de distinguer un texte en anglais d'un texte en français. Qu'en pensez-vous ? Testez votre méthode et présentez les résultats.

On peut utilisé la méthode du χ^2. Pour cela il faut calculer les valeurs des χ^2 pour le français et pour l'anglais afin de mesurer la différence entre les fréquences des deux langues. Le texte est ensuite classé dans la langue ayant obtenu la plus grande valeur de χ^2. Cet algorithme fonctionne pour des textes suffisament long (min 50 caractères). Si la différence entre les χ^2 des deux langues est trop faible, l'algorithme retourne indéterminé.

Résumé:

1. Décompter chaque lettre présente dans le texte
2. Calculer la fréquence de chaque lettre dans le texte
3. Calculer le χ^2 pour chacune des langues (avec chacune son référentiel de fréquence)
4. Comparer les valeurs des χ^2, la valeur la plus haute indique la langue.

(exemple dans le fichier quelle_langue.py)

8. Quelles étaient les clef et les textes clairs correspondants aux textes chiffrés dans les fichiers vigenere.txt et vigenere_improved.txt ?

Pour le texte *vigenere.txt*

- **Clé:** asterixobelix
- **Texte clair:** voussavezmoijenecroispasquilyaitdebonneoudemauvaisesituationmoisijedevaisresumermavieaujourdhui avecvousjediraisquecestdaborddesrencontresdesgensquimonttendulamainpeutetreaunmomentoujene pouvaispasoujetaisseulchezmoietcestassez curieuxdesedirequeleshasardslesrencontresforgentunedestineeparcequequandonalegoutdelachosequandonalegoutdelachosebienfaitele beaugesteparfoisonnetrouvepaslinterlocuteurenfacejediraislemiroirquivousaideaavanceralorscanestpasmoncascommejedisais lapuisquemoiaucontrairejaipuetjedismercialaviejeluidismercijechantelavie jedanselaviejenesuisquamouretfinalementquanddesgensmedisentmaiscommentfaistupouravoircettehumanitejeleurrepondstressimplement quecestcegoutdelamourcegoutdoncquimapousseaujourdhuiaentreprendreuneconstructionmecaniquemaisdemainquisaitpeutetresimplement amemettreauservicedelacommunauteafaireledonledondesoi

Pour le texte *vigenere_improved.txt*

- **Clé:** justeleblanc
- **Texte clair:** jeprendslaccentbelgenonalloopourraisjeparleramonsieurleblancjusteunefoecestmoibonsoermonsieurleblancgeorgevanbrugel alappareiljevousappelleparcequejesuisproducteurnestcepasjarrivedebelgiqueunefoeetjesuistresinteresseparvotreromanvotreroman lepetitchevaldemanegelepetitchevaldemanegeetjaimeraisdiscuter lachatdesdroepourlecinemecestuneblagueouquoinonnonpasdutout pourquoiuneblegue